

Tackling the tensions in evaluating capacity strengthening for health research in low- and middle-income countries

Imelda Bates,^{1*} Alan Boyd,² Garry Aslanyan³ and Donald C Cole⁴

¹Liverpool School of Tropical Medicine, Liverpool, UK, ²Manchester Business School, Booth St W, Manchester, Greater Manchester M15 6PB, UK, ³Tropical Disease Research, World Health Organization, 20, Avenue Appia, CH-1211 Geneva 27, Switzerland and ⁴Dalla Lana School of Public Health University of Toronto, 155 College St. Toronto, ON M5T3M7, Canada

*Corresponding author. Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK. E-mail: ibates@liv.ac.uk

Accepted 13 February 2014

Strengthening research capacity in low- and middle-income countries is one of the most effective ways of advancing their health and development but the complexity and heterogeneity of health research capacity strengthening (RCS) initiatives means it is difficult to evaluate their effectiveness. Our study aimed to enhance understanding about these difficulties and to make recommendations about how to make health RCS evaluations more effective. Through discussions and surveys of health RCS funders, including the ESSENCE on Health Research initiative, we identified themes that were important to health RCS funders and used these to guide a systematic analysis of their evaluation reports. Eighteen reports, produced between 2000 and 2013, representing 12 evaluations, were purposefully selected from 54 reports provided by the funders to provide maximum variety. Text from the reports was extracted independently by two authors against a pre-designed framework. Information about the health RCS approaches, tensions and suggested solutions was re-constructed into a narrative. Throughout the process contacts in the health RCS funder agencies were involved in helping us to validate and interpret our results. The focus of the health RCS evaluations ranged from individuals and institutions to national, regional and global levels. Our analysis identified tensions around how much stakeholders should participate in an evaluation, the appropriate balance between measuring and learning and between a focus on short-term processes vs longer-term impact and sustainability. Suggested solutions to these tensions included early and ongoing stakeholder engagement in planning and evaluating health RCS, modelling of impact pathways and rapid assimilation of lessons learned for continuous improvement of decision making and programming. The use of developmental approaches could improve health RCS evaluations by addressing common tensions and promoting sustainability. Sharing learning about how to do robust and useful health RCS evaluations should happen alongside, not after, health RCS efforts.

Keywords Low- and middle-income countries, health, research, capacity strengthening

KEY MESSAGES

- Tensions in conducting evaluation of health research capacity strengthening (RCS) centre around how much stakeholders should participate in an evaluation, the appropriate balance between measuring and learning and balancing measures of short-term processes against longer-term impact and sustainability.
- Ways of avoiding these tensions and therefore of making these evaluations more effective include early and ongoing stakeholder engagement in planning and evaluating health RCS, modelling of impact pathways and rapid assimilation of lessons learned for continuous improvement of decision making and programming.
- Sharing learning about how to do robust and useful health RCS evaluations should happen alongside the health RCS efforts, not after they have finished.

Introduction

Strengthening research capacities in low- and middle-income countries (LMICs) is one of the most effective ways of advancing their health and development (van Velzen *et al.* 2009). Promising approaches in strengthening health research capacity (RCS) in LMICs have been documented (Ghaffar *et al.* 2008; Bennett *et al.* 2010) but demonstrating the effectiveness of the significant investments that have been made in health RCS is challenging (Whitworth *et al.* 2008; Breman *et al.* 2011). In 2008–09, we reviewed 593 published peer-reviewed papers concerning health RCS evaluations but discovered only 4 primary, longitudinal evaluations of health RCS interventions in low- or middle-income countries; the rest were predominantly descriptions of individual programmes or proposed frameworks (Cole DC, Boyd A, Aslanyan G, Bates I, under review). The lack of robust evidence on what works in conducting health RCS evaluations has been discussed at several high-profile international meetings over the last 2 years such as those organized by the Association of Commonwealth Universities (2012),¹ London International Development Centre (2012),² Forum 2012³ and The Alliance for Health Policy and Systems Research (2012).⁴

The monitoring and evaluation of programmes that aim to strengthen capacity to generate health research [i.e. health research capacity strengthening (RCS) programmes] is widely considered to be problematic because of the diversity of contexts in which RCS takes place, difficulties in attribution, the long timescales needed to demonstrate sustainability and the limited availability of LMIC-based evaluators with appropriate skills (HCSTC 2012). Recognizing these difficulties, a group of health RCS funding agencies came together in 2008 as the ESSENCE on Health Research initiative⁵ and identified monitoring and evaluation as one of the areas for potential harmonization of good practices. They produced a planning, monitoring and evaluation (PM&E) framework as a guide for their members and their grantees (ESSENCE on Health Research 2011) and encouraged sharing of lessons about health RCS evaluations.

A better understanding of the strengths and limitations of the various approaches used to evaluate health RCS, the assumptions that underpin the various approaches, and the tensions and challenges that exist, would help all those involved in health RCS to make evaluations more useful. We therefore initiated a research project in collaboration with key contacts in health RCS funding agencies including the ESSENCE group. We asked funders what they sought in evaluations, examined

existing frameworks for and reports on evaluations and more broadly canvassed stakeholders regarding their experience of using frameworks in evaluating health RCS. We synthesized our findings under three categories: the frameworks used to guide RCS evaluations, evaluation design and indicators used and the processes and tensions involved in health RCS evaluations. This latter analysis is the focus of this article in which we aim to understand the rationale behind the choice of evaluation approaches and to highlight and critique solutions proposed in the evaluation reports we examined.

Methods

Methods of stakeholder engagement, collection of relevant documents and evaluation selection (Lincoln and Guba 1985; Maessen 2005; Keown *et al.* 2008; Smith *et al.* 2009; Saini and Shlonsky 2012) are described in Box 1. We identified themes in health RCS that were considered important through discussions with key funder contacts and from their organizations' strategy documents. We conducted a systematic analysis (Pope *et al.* 2000) on the evaluation reports by extracting text about the health RCS approach, tensions encountered during the evaluation and suggested solutions, into a pre-designed framework. We mapped the text against the themes in health RCS that had been identified as important to funders. Two authors independently extracted text from each evaluation report provided by the funders. Consistency was checked across reports, and we resolved discrepancies through discussion between authors. We reconstructed the aggregated text as a narrative, highlighting the rationale for and against particular approaches within each theme and summarizing reported ways of addressing tensions between the various approaches. We stopped analysis when no new insights emerged from analysis of additional reports.

Scrutiny of funders' organizational strategy documents and discussions with contacts in the funding agencies helped us to identify themes within health RCS that they currently consider to be important. This information, combined with feedback on our interim findings by our funder contacts at several stages of the project, helped us to focus our analysis and to validate and interpret our results.

Results

The eighteen reports (Pederson *et al.* 2011; Erlandsson and Gunnarsson 2005; Zuckerman *et al.* 2006; HERA 2007a,b; Agepong *et al.* 2008; Peebles and Sachdeva 2008; Sachdeva

Box 1. Methods for analysis of processes and tensions in evaluations of health RCS (August 2011 to March 2013)

Stakeholder engagement:

- partnership with ESSENCE on Health Research initiative—research team member and consultations with steering committee, in keeping with reviews involving stakeholders (Keown *et al.* 2008; Smith *et al.* 2009);
- survey to identify funding agencies interested in participation (11/31 agreed);
- telephone discussions with key funder contacts and staff;
- meetings and workshop at the Global Health Forum 2012 (<http://www.forum2012.org/>);
- comments and suggestions helped analysis and interpretation in an iterative fashion, producing an integrative meta-synthesis (Saini and Shlonsky 2012).

Collection of English language documents, all but one published on funder agency websites:

- Reports of health RCS evaluations, 2000 forward. Although numerous activity and financial monitoring reports regularly provided by grantees but confidential and not regarded as evaluations by our funder contacts. Yield was 54 reports agreed relevant by pairs of reviewers.
- Maximum variety sampling (Lincoln and Guba 1985) of evaluations based on reading and summaries of each report by pairs of reviewers. Eighteen reports (12 evaluations) purposively selected as detailed (Maessen 2005), a focus on health, of a variety of programmes, from a range of funders, in a range of countries, and with diverse evaluation approaches.
- Health RCS evaluation frameworks and evaluation policy documents collected from references of documents and through websites.

Online survey of health RCS funders, implementers and evaluators about health RCS evaluation frameworks (response rate 25/48, 52%), to supplement those above and those identified in a systematic review (Cole *et al.* 2012).

Formal approval of ethics protocol (#26837, University of Toronto Health Sciences Research Ethics Board).

and Peebles 2008; Sachdeva *et al.* 2008a,b; Day *et al.* 2009; Srivastava *et al.* 2009; van Velzen *et al.* 2009; Vullings and Meijer 2009; Mills *et al.* 2010; Minja and Nsanjabana 2010; Podems 2010; Ransom *et al.* 2010) describe 12 different evaluations which spanned individual, institutional, national, regional and global levels (see Table 1). The important health RCS evaluation themes identified during the initial phase of the project were participation, impact, learning and timescale. Examples of the tensions encountered during health RCS evaluations are provided in Box 2 and suggested solutions mentioned in the evaluation reports are summarized below.

Participation

Tension: *to what extent should funding recipients participate in health RCS evaluations?*

Tensions around the degree to which funding recipients should be involved in the evaluation of their own health RCS efforts were reflected in many of the reports. Reasons given for promoting an external, non-participative approach included that it was better for accountability, for assessing value for money and for quick results. The lack of expertise among funding recipients' in setting testable goals and measurable targets, or in evaluation techniques was also cited as a reason not to choose participatory evaluations. In contrast, reasons why recipients should participate in the evaluation were that it promoted ownership, learning and implementation of recommendations. Compared to external evaluators, funding recipients were perceived to have better in-depth knowledge about the project, the stakeholders and the context. Such knowledge was considered important for problem solving and sustainability.

Engaging stakeholders (e.g. service users, community members, health practitioners and policy makers) in the evaluation was considered to be helpful for setting realistic goals, ensuring alignment with local priorities and for addressing resource issues. However, it was acknowledged that extensive participation required resources (e.g. time and expenses) and infrastructure (e.g. communications and networks) which would need to be budgeted for and planned. Contrasting views about how to avoid bias in evaluations and for ensuring transparency were evident; for example, one report advocated the use of an 'external contract organization', while another preferred stakeholder participation (Agyepong *et al.* 2008; van Velzen *et al.* 2009).

The extent of recipients' participation in evaluations varied between reports. This may be partly explained by the fact that participation was sometimes interpreted as 'access' or 'information exchange' (van Velzen *et al.* 2009). Although many reports stated that participation was desirable, there was limited evidence of significant participation by recipients in practice. Four indicated some recipient participation in the evaluation (Erlandsson and Gunnarsson 2005; HERA 2007a,b; Agyepong *et al.* 2008) and another described consultations with stakeholders about project implementation (Srivastava *et al.* 2009). Two reports indicated that although participation in evaluation was encouraged, it was limited by lack of funding (Erlandsson and Gunnarsson 2005; Agyepong *et al.* 2008). Another, while recognizing the benefits of building partnerships among beneficiaries, advocated outsourcing project monitoring (van Velzen *et al.* 2009).

Suggested solutions mentioned in the evaluation reports

A desire to combine the perceived higher quality of external evaluations with the learning, ownership and sustainability associated with participatory evaluations was evident in several reports. In an attempt to increase the quality of participation in health RCS evaluations some funders had supported training in evaluation for recipients. More interaction between funders, evaluators and funding recipients from the outset was suggested to allow time to share knowledge, develop trust and institutionalize involvement in evaluation (Agyepong *et al.* 2008; Vullings and Meijer 2009; Podems 2010).

Table 1 Characteristics of funders' evaluations of health RCS

Evaluations	Health RCS characteristics		Relation of evaluator to funder	Evaluation characteristics	
	Project, programme(s), organization	Period (duration) covered by the evaluation		Timing	Main approaches/methods
Funder Lead author of report (year of publication)					
ACU-CSC Ransom <i>et al.</i> (2010), Day <i>et al.</i> (2009)	CS programme—with streams, health sector	1960+ (48 years)	Funder staff	Periodic review	Analysis of existing award data, alumni evaluation survey, 15 case studies and five telephone interviews of selected scholarship recipients. Impact assessment.
Carnegie Podems (2010)	RCS initiative with networks	2008–10 (2 years)	Contract evaluation organization	Mid-term	Desk review of initiative and network documents, interviews and focus groups with stakeholders (key staff and students within each network).
Danida Health Research for Action (HERA) (2007a,b)	HR programmes of which health RCS a part	1997–2006 (10 years)	Contracted evaluation organization	Periodic review	Components were: (a) country reports with visits; (b) desk study review of projects; (c) institutional questionnaires for Danish research groups; (d) 'internal' (Danish organizations) individual staff questionnaires; (e) 'external' (non-Danish other High Income Country funder) questionnaires and interviews; (f) literature review of publications supported; (g) evaluation document analysis and (h) health-related project database analysis.
DHID Mills <i>et al.</i> (2010)	Project-HR council	2008–10 (2 years)	External programme evaluation team	Mid-term review	Desk review of organizational, programme and project documentation; site visit with interviews of stakeholders, beneficiaries, non-beneficiaries, funders and secretariat; in-depth case studies of selected grantees and their institutions and evaluation of the grants selection process.
EDCTP van Velzen <i>et al.</i> (2009)	HR partnership	2007–09 (2 years)	Independent external panel	Periodic review	Documentation analysis, meetings—discussions and interviews with organizational representatives, questionnaire survey of researchers, site visit, conference attendance and country case study.
IDRC Sachdeva <i>et al.</i> (2008a,b), Sachdeva and Peebles (2008), Peebles and Sachdeva (2008)	HR programme with projects	Roughly 2001–08 (7 years)	Contracted evaluation team	Special review	Conducted a gender audit at three levels—institutional, programmatic and project (review of 15 projects)—through documentation review; search of guidelines and strategies of other organizations working on policy, health and gender issues; review of a previous internal gender survey; gender questionnaire to assess capacity development needs and individual interviews with funder staff.
NIH-FIC (1) Zuckerman <i>et al.</i> (2006)	Health RCS Programme	1992–2003 (11 years)	Contract evaluators	Periodic review	Outcome evaluation using NIH-FIC evaluation framework and Fogarty International Research Collaboration Award logic model. Administrative data collection and review, interviews with programme stakeholders, census surveys of the US principal investigators and international research collaborators, bibliometric analysis of publications and site visits.

(continued)

Table 1 Continued

Evaluations		Health RCS characteristics		Relation of evaluator to funder	Evaluation characteristics	
Funder	Lead author of report (year of publication)	Project, programme(s), organization	Period (duration) covered by the evaluation		Timing	Main approaches/methods
NIH-FIC (2)	Strivastava <i>et al.</i> (2009)	Health RCS Programme	2002–08 (6 years)	Contract evaluation team	Mid-term review	Programme implementation and preliminary outcomes. Data collection methods included two online surveys (Global Research Initiative Program (GRIP) awardees, unsuccessful applicants with scored applications). Supplementary data from administrative sources and databases, MEDLINE, and from interviews with US-based mentors, FIC staff members and programme partners.
NWO/WOTRO	Agyepong <i>et al.</i> (2008), Vullings and Meijer (2009)	Health RCS and HR programmes	2005–08 (4 years)	[1a] Committee of three experts and two secretariat members [1b] Contract evaluators	Mid-term review	[1a]Background document review, discussions with programme co-ordinators, site visits with interviews, formulate recommendations and discuss with Programme Committee [1b] Not specified but included: programme document review, programme logic construction, projects' progress reports analysis and stakeholder interviews.
SIDA	Erlandsson (2005)	Linked health RCS project funding (three routes)	1999–2005 (6 years)	Contract evaluators	Mid-term for re-formulation	Emailed questionnaires to institutions, individuals and graduates. Interviews during site visits and evaluation seminar at main site.
TDR-WHO	Minja and Nsanzabana (2010)	Organization's entire set of health RCS programmes	2000–08 (9 years)	Contracted institute evaluation team	Periodic review	Questionnaires (individuals, research groups and institutions), selected in-depth interviews, institutional site visits with stakeholder semi-structured interviews.
Wellcome Trust	Pederson <i>et al.</i> (2011)	Health RCS Project—Consortium	2009–11 (2 years)	Contract evaluation organization	Mid-term (second annual)	Real-time, monitoring and evaluation with mutually agreed framework of qualitative and quantitative indicators. Analysis in the light of all consortia within the programme of which this project is a part.

CS, capacity strengthening; HR, health research; ACU-CSC, Association of Commonwealth Universities-Commonwealth Scholarship Commission; Carnegie, Carnegie Corporation of New York through Science Initiative Group; Danida, Development co-operation activity, Ministry of Foreign Affairs (Denmark); DFID, Department for International Development (UK); EDCTP, European and Developing Countries Clinical Trials Partnership; IDRC, International Development Research Centre (Canada); NIH-FIC, National Institutes of Health-Fogarty International Center (USA); NWO/WOTRO, Science for Global Development, Netherlands Organization for Scientific Research; SIDA, Swedish International Development Agency; TDR-WHO, Tropical Disease Research-World Health Organization.

Box 2. Illustrative examples of tensions in health research funders' reports**Participation**

Setting: Sub-Saharan Africa. *Evaluation focus:* individuals, institutions, networks.

Example: Individuals involved in the project...are more likely to interpret the materials and findings in ways that are understood by the majority of the project members. They are part of the socio-cultural context...that is not necessarily shared by the evaluators. This participant evaluation model also facilitates the process of change that is often a necessary next step for most projects (Erlandsson and Gunnarsson 2005).

Setting: Primarily Africa, also Asia and inter-regional. *Evaluation focus:* individuals, some institutions, occasional national systems.

Example: Given the relatively short time horizon and the overall resource frame for the current review, it was not possible to assess a representative sample of research projects through field visits. Also, for practical reasons and time constraints, more consultations took place with North than with South stakeholders (HERA 2007a).

Impact

Setting: Sub-Saharan Africa. *Evaluation focus:* individuals, institutions, networks.

Example: While data indicate the usefulness of networks, it is too early in the programme' implementation to determine the concrete results (Podems 2010).

Setting: Multiple regions. Individual scientists and their institutions.

Example: Longer-term follow-up and more rigorous evaluation design would be needed to assess the outcomes and impacts of GRIP (Srivastava *et al.* 2009).

Learning

Setting: Africa. *Evaluation focus:* sub-regional networks.

Example: An explicit objective was to draw out any lessons to be learned and to formulate recommendations for future initiatives. Such recommendations were produced. The evaluation also considered the previous review (there were some common members in the evaluation teams) and the "internal" review. It made a point of endorsing and restating previous recommendations, and observing that some "have still not been entirely fulfilled" (van Velzen *et al.* 2009).

Sustainability

Setting: Primarily Africa, also Asia and inter-regional. *Evaluation focus:* individuals, some institutions, occasional national systems.

Example: Moreover, there was insufficient planning for the sustainability of institutional collaborative arrangements, and no phasing-out plans were apparent (HERA 2007a).

The TOR [terms of reference] are broad in scope and that the time available for the review was short...some of the work would be less detailed and some aspects of the TOR not addressed in a comprehensive manner in view of the time constraint (HERA 2007a).

Setting: Sub-Saharan Africa. *Evaluation focus:* individuals, institutions, environment and networks.

Example: The first dilemma is the short-term project-period which asks for clear timelines, targets and milestones while at the same time sustainability and local ownership is important. The latter requires trust-building which is a long-term process (Agyepong *et al.* 2008).

Impact***Tension: should evaluation efforts focus on understanding processes or measuring impact?***

Tensions between the desire for only a few common measurable, reliable, indicators and the need to evaluate project impact arose in many of the reports. There were also some disparities about whether health RCS 'impact' referred to health outcomes, research capacity or both. Retrospective

evaluations, which predominated in the reports, were thought to provide nothing more than an 'educated guess' with the absence of baseline data and pre-determined indicators threatening the validity of the findings (van Velzen *et al.* 2009). A trade-off was therefore apparent between the need for evaluations that were valid despite the complexity and uniqueness of projects, and the constraints of limited time and resources.

The need to use both quantitative and qualitative indicators specifically designed for each project made it difficult to identify a set of easily measurable indicators that provided meaningful information about health RCS impact. In practice, the reports tended to focus on processes thought to lead to better research capacity rather than measuring impact directly. Evaluations of impact on individuals' research esteem (e.g. researchers' publications, invitations to speak and job offers) (Srivastava *et al.* 2009; Minja and Nsanzabana 2010) or institutional research capacity (Erlandsson and Gunnarsson 2005; Zuckerman *et al.* 2006; HERA 2007a,b; Srivastava *et al.* 2009) were more common than evaluations of the use of research results to inform health policies and programmes.

Suggested solutions mentioned in the evaluation reports

A comprehensive, prospective system for health RCS evaluation was suggested, in which long-term impact across different levels is considered throughout the whole project cycle (Srivastava *et al.* 2009; Minja and Nsanzabana 2010; Podems 2010) using clear conceptual frameworks, multiple data sources and valid standards to enhance quality. Specific capacity strengthening plans and modelling of the expected health RCS impact over time, for example, using outcome mapping and logical frameworks was considered a sound basis for selecting indicators and setting targets (van Velzen *et al.* 2009). Collaborations between funders and with other stakeholders were considered helpful for supporting broader, long-term evaluations and strategically targeting impact evaluations.

Learning

Tension: is the purpose of health RCS evaluations to demonstrate accountability or to enhance knowledge?

Enhancing knowledge about how to make health RCS efforts more effective was recognized as an important function of health RCS evaluations (Vullings and Meijer 2009; Podems 2010). Emphasis was placed on 'learning by doing' and knowledge that was relevant beyond the particular project being evaluated. Reports recommended that the lessons should be documented systematically, and shared between funders and between projects. It was recognized that funders faced tensions between their need to show accountability and value for money, and the extent to which they should invest in facilitating and sharing learning. It was also suggested that time spent by researchers on documenting and using learning may divert them from their primary research and health RCS activities unless the timescales and funding for the research were flexible enough to avoid such tradeoffs (Vullings and Meijer 2009). Two reports highlighted that the evaluation itself had enabled learning, evidenced by actions of stakeholders and funders in response to the evaluation findings (Erlandsson and Gunnarsson 2005; Agyepong *et al.* 2008).

Suggested solutions mentioned in the evaluation reports

No report explicitly considered how funders could balance the demand for accountability with the need to maximize learning opportunities, though there were many suggestions about how learning could be promoted within projects. These included effective communications between funders, evaluators and recipients within and between projects and institutions.

Face-to-face dialogue to discuss progress from early on in the project was found to be particularly useful (Zuckerman *et al.* 2006; Agyepong *et al.* 2008; Vullings and Meijer 2009). Benchmarking and sharing of experiences through, for example, networking and exchange visits, were reported to enhance informal learning about topics such as governance (Agyepong *et al.* 2008; van Velzen *et al.* 2009). Reports indicated that flexible institutional systems able to rapidly assimilate findings from the evaluation process into decision making and programming, as well as regular participatory evaluation or self-assessment could all contribute to lesson learning for continuous improvement.

Timescale

Tension: how can short-term funding be reconciled with the need for long-term sustainability?

Tensions between short-term funding to conduct and evaluate projects, the 5–10 years needed to develop sustainable capacity (van Velzen *et al.* 2009; Bates *et al.* 2011) and the two decades needed to show impact (Vogel 2011) were apparent in several reports. Within the reports sustainability seemed to be represented by heterogeneous concepts such as avoiding early collapse of a project, continuation of a project without dependence on the original funder and durable health research capacity emerging out of a project. Only one report (Zuckerman *et al.* 2006) explicitly documented their criteria for assessing sustainability. Several reports evaluated how sustainability of projects or associated capacity development had been facilitated (Erlandsson and Gunnarsson 2005; Zuckerman *et al.* 2006; HERA 2007a,b; Agyepong *et al.* 2008; Vullings and Meijer 2009) though two focused solely on financial sustainability (HERA 2007a,b). Some reports (HERA 2007a,b; Agyepong *et al.* 2008; Srivastava *et al.* 2009; Minja and Nsanzabana 2010; Podems 2010) considered how the health RCS evaluation related to the maturity of the project, and three specifically examined whether funders' expectations were realistic within the project's timescale (Srivastava *et al.* 2009; Minja and Nsanzabana 2010; Podems 2010).

Suggested solutions mentioned in the evaluation reports

To resolve the tension between the time available to conduct the evaluation and the need to evaluate sustainability, reports highlighted actions that could be taken at different system levels to promote sustainability and which could be evaluated within the project timeframe. Suggestions in the reports included recipients planning for sustainability from the start (HERA 2007a,b; Vullings and Meijer 2009), focusing health RCS around specific research questions (HERA 2007a,b), obtaining matched funding (van Velzen *et al.* 2009), building on existing partnerships (Zuckerman *et al.* 2006; Vullings and Meijer 2009) and choosing partners who were influential or had a track record of attracting funding (HERA 2007a,b; Agyepong *et al.* 2008; Podems 2010). Actions for institutions to promote sustainability included obtaining international accreditation (Vullings and Meijer 2009) and organizational integration of health RCS initiatives (Erlandsson and Gunnarsson 2005; HERA 2007a,b; Agyepong *et al.* 2008; Vullings and Meijer 2009). Funders could provide specific health RCS funding (Vullings and Meijer 2009) and engage in

strategic partnerships (e.g. national programmes, international organizations) to share the load of supporting long-term initiatives for sustainability (van Velzen *et al.* 2009). Evaluations themselves were considered to have the potential to promote sustainability (Erlandsson and Gunnarsson 2005; Agyepong *et al.* 2008; Vullings and Meijer 2009) and prevent ‘fade’ of neglected components such as gender balance (Peebles and Sachdeva 2008; Sachdeva *et al.* 2008a,b). Reports stressed the importance of flexible, early and regular process evaluations, and of encouraging local ownership of all project stages, including evaluation, for promoting sustainability (HERA 2007a,b; Agyepong *et al.* 2008).

Discussion

We have presented four key tensions, which we identified in the funders’ health RCS evaluations, as if they are independent of each other, but in practice they are often interlinked. The choice about whether to focus on evaluating project processes or the more difficult-to-measure ‘impact’ is closely tied to issues of short-term project funding and aspirations of creating sustainable change. The choice about whether to have substantial participation of stakeholders in evaluations or to use exclusively independent evaluators is strongly influenced by whether the evaluation is to be used primarily for accountability or for learning. We have taken into account the interdependence of potential tensions in synthesizing evidence from our study and published literature to provide a holistic perspective about how evaluations of health RCS could be made more effective. Our study focused primarily on issues of importance to funders and therefore on funder-led priorities and programmes. Issues, solutions and evaluation criteria may therefore differ from those considered important by, for example, funding recipients or policy makers.

Defining the explicit purpose and intended use of health RCS evaluations could alleviate tensions that hinder their effectiveness

Our findings demonstrate that the scarcity of information about the intended use of the evaluation and the rationale for the approach used, could contribute to a mismatch between the funders’, recipients’ and evaluators’ expectations of what the evaluation should achieve. Evaluations that are ‘utilization-focused’ (Patton 1994) have an explicit purpose and are designed to meet the primary users’ requirements; they may therefore be more likely than traditional evaluations to provide good value for money. We recognize that funders have internal and external perspectives and the evaluation reports we analysed may have been focused primarily on an internal audience. Specifying the intended use of the evaluation helps to guide choices about its purpose (formative, summative, developmental), data type (quantitative, qualitative, mixed), design (naturalistic, experimental) and focus (process, outcomes, costs, cost benefit).

Evaluations as a learning tool

Research funders are being urged by governments to ensure that capacity building initiatives become self-sustaining

(HCSTC 2012). This is more likely to be achieved with a developmental type of evaluation than with traditional summative evaluation approaches. From the reports we analysed it was apparent that evaluators were often aware that theory-driven evaluative thinking can lead to more rigorous and useful evaluations (Sridharan and Nakaima 2011). However, they had insufficient time and resources to prospectively incorporate theory-informed indicators of impact and sustainability into the evaluation. Our results show that this resulted in missed opportunities to enhance knowledge among funders and funding recipients about how to improve PM&E of health RCS programmes Boyd *et al.* 2013. Very few reports (Minja and Nsanzabana 2010) reviewed progress in health RCS against previous recommendations. It is possible that we may have underestimated the extent of such ‘developmental’ learning because we did not track individual project reports over time, due partly to issues over confidentiality and access. Developmental evaluations are especially useful where the purpose is learning, innovation and change rather than external accountability. Ideally developmental evaluations, and the related utilization-focused evaluations, should have a 20-year flexible perspective and a focus of accountability towards non-funder stakeholders in LMICs (Patton 2011). Thorough negotiations with key stakeholders and transparent decision making are important to avoid criticisms about credibility. While we recognize that funders may be constrained by their own institutions’ monitoring and evaluation requirements, for example, logical frameworks (Hovland 2007), incorporating developmental principles into future health RCS evaluations may help to alleviate some of the tensions surrounding the extent of learning and recipients’ participation in evaluations.

Engaging research funders, recipients and evaluators in describing the intervention pathway for health RCS to produce useful evaluations

A key message that emerged from our analysis is that close and regular dialogue among all those involved in the health RCS evaluation is essential to achieve a jointly agreed purpose, to maintain engagement and momentum and to provide valid and balanced findings. Capacity strengthening is a complex process with a long intervention pathway. Applying a theory of change—a description of the relationships between activities, outputs and outcomes—can help to define a pathway and inform the evaluation by delineating impact trajectories and fostering generalizability assessment, evaluation rigour and policy and practice influence (Sridharan *et al.* 2006; Pawson and Sridharan 2009; Kubisch *et al.* 2010). Evaluators are well positioned to facilitate discussions about this pathway among stakeholders and to promote empowerment by enhancing stakeholders’ skills in independent problem solving and decision making. Typically ‘empowerment evaluations’ are characterized by recipients’ efforts to gain control of programmes, to obtain resources and to critically understand the social environment (Fetterman 2010). Cadres of stakeholders need specific skills to help them effectively fulfill their role in health RCS evaluations: funders to commission, procure, participate in and use evaluations; recipients to participate in designing monitoring and evaluation processes and in generating data; evaluators to combine technical, organizational and

Table 2 Recommendations to avoid tensions in health RCS evaluations synthesized from evaluation reports and categorized by theme**Participation**

Funders and recipients (and evaluators) will design a better project, learn more, have more ownership of recommendations and be more likely to implement them if they are involved closely in developing and formulating the CS evaluations.

Impact

A comprehensive and planned approach is required to demonstrate the value that health RCS brings. Impact and value for money should be considered throughout a project, from design of initiatives through to influence at systems, institutional and political levels.

Learning

Lessons learned about health RCS should be systematically documented and shared within and between funders and between projects, and used to inform organizational decision making and programme planning.

Timescale

Sustainable capacity development has a long time frame and needs long- and short-term monitoring and evaluation (M&E). M&E activities should be beneficial, feasible (e.g. adequate resources), avoid unnecessary administrative burden and be used for learning. Research timescales need to be flexible if effective CS is to be achieved alongside a research project.

interpersonal skills, with an understanding of the complexity of health RCS and knowledge of a range of evaluation processes and instruments (EU 2008).

A way forward: a step-change in learning and sharing to accelerate the effectiveness of health RCS evaluations

Unintended negative consequences may arise from health RCS initiatives if they are not carefully planned (Vasquez *et al.* 2013). The impact and nature of RCS may be influenced by political and other constraints on the funding organizations and by the difficulties researchers face in effectively collaborating in capacity strengthening for research-to-policy translation.

Awareness of the need to develop robust evaluation approaches, particularly ones that capture the longer-term post-programme benefits of health RCS, is increasing. Carefully designed multi-programme comparative evaluation that takes account of differences in programmes and their contexts would be helpful in determining the most cost effective and sustainable RCS models for given contexts.

Our recommendations for accelerating the effectiveness of health RCS evaluation are synthesized in Table 2. A critical recommendation is the need to actively involve funding recipients and other stakeholders in all stages of the evaluation process. This could enable detection and correction of problems throughout the project lifespan, make the decisions underlying evaluations more visible and encourage sharing and utilization of the evaluation results (HCSTC 2012). The timing of this knowledge sharing is important and should be rapid to avoid duplicating ineffective approaches thereby wasting time and resources. It must happen alongside, not after, health RCS efforts, to shorten the 6-year time lag it takes to incorporate lessons from end-of-project evaluations into commissions for new research programmes (HCSTC 2012).

Over the last few years research funders have made significant efforts to learn from each other about health RCS evaluations in LMICs through initiatives such as ESSENCE on Health Research and the Collaborative on Development Sciences (UKCDS) (ESSENCE on Health Research 2011; Vogel 2011). By making explicit the tensions we have uncovered in funders' health RCS reports, we hope to promote and contribute to the critical debate between such coalitions and the research community, and accelerate progress in developing robust and useful evaluations of health RCS.

Acknowledgements

We would like to thank all those who participated in the interviews and survey as well as Ritz Kakuma who participated in project formulation, Maniola Sejrani for initial project work, Quenby Mahood for help with document selection and tracking and Dan-Bi Cho for her contribution to data extraction. We thank the reviewers for insightful comments.

Funding

This work was supported by the Canadian Institute of Health Research: IIM-111606.

Conflict of interest statement: Garry Aslanyan is part of the secretariat of the ESSENCE on Health Research initiative. The other authors have no competing interests.

Endnotes

- ¹ Association of Commonwealth Universities Members Day Towards a 'Global' Institution—6 July 2012, London 2012. <https://www.acu.ac.uk/>, accessed 23 May 2013.
- ² London International Development Centre Measuring Impact of Higher Education for Development, 19–20 March 2012, accessed 23 May 2013.
- ³ Council for Health Research for Development and The Global Forum for Health Research. Cape Town 2012. <http://www.forum2012.org/>, accessed 23 May 2013.
- ⁴ Global Symposium on Health Systems Research. <http://www.who.int/alliance-hpsr/hsr-symposium/en/>, accessed 23 May 2013.
- ⁵ <http://www.who.int/tdr/partnerships/initiatives/essence/en/>, accessed 23 May 2013.

References

- Agyepong I, Kitua A, Borleffs J. 2008. *Mid-term Evaluation Report, NACCAP Programmes First Call: INTERACT*. The Hague: NACCAP.
- Bates I, Taegtmeier M, Squire SB *et al.* 2011. Indicators of sustainable capacity building for health research: analysis of four African case studies. *Health Research Policy and Systems* **9**: 1–9.
- Bennett S, Paina L, Kim C *et al.* 2010. What must be done to enhance capacity for Health Systems Research? In: World Health Organization (ed). *Background papers commissioned by the Symposium Secretariat for the First Global Symposium on Health Systems Research*, November 16–19, 2010, Montreux, Switzerland. <http://www.rockefellerfoundation.org/uploads/files/c4ccb675-f6f8-47de-8552-e032d4c3fc20.pdf>, accessed 3 March 2014.

- Boyd A, Cole DC, Cho DB, Aslanyan G, Bates I. 2013. A comparison of frameworks for evaluating health research capacity strengthening: what can we learn? *Health Research Policy and Systems* **11**: 46.
- Bremar JG, Bridbord K, Kupfer LE, Glass RI. 2011. Global health: the Fogarty International Center, National Institutes of Health: vision and mission, programs, and accomplishments. *Infectious Disease Clinics of North America* **25**: 511–36.
- Cole DC, Kakuma R, Fonn S *et al.* 2012. Evaluations of health research capacity development: a review of the evidence. *American Journal of Tropical Medicine and Hygiene* **87**(5 Suppl. 1):801.
- Day R, Stackhouse J, Geddes N. 2009. *Evaluating Commonwealth Scholarships in the United Kingdom: Assessing Impact in Key Priority Areas*. London: Commonwealth Scholarship Commission in the UK. <http://cscuk.dfid.gov.uk/wp-content/uploads/2011/03/evaluation-impact-key-report.pdf>, accessed 26 April 2013.
- Erlundsson B, Gunnarsson V. 2005. *Evaluation of HEPNet in SSA*. Stockholm: Sida. <http://www.sida.se/Publications/Import/pdf/sv/Evaluation-of-HEPNet-in-SSA.pdf>, accessed 26 April 2013.
- ESSENCE on Health Research. 2011. *Planning, Monitoring and Evaluation Framework for Capacity Strengthening in Health Research*. Geneva: ESSENCE on Health Research. <http://www.who.int/tdr/publications/non-tdr-publications/essence-framework/en/index.html>, accessed 26 April 2013.
- European Union. 2008. *Developing Evaluation Capacity: Final Report on the Framework to Analyze the Development of Evaluation Capacity in the EU Member States*. 2nd edn. Luxembourg: European Union. http://ec.europa.eu/regional_policy/sources/docgener/evaluation/pdf/report_integrated_2007.pdf, accessed 26 April 2013.
- Fetterman D. 2010. *Empowerment Evaluation: Collaboration, Action Research, and a Case Example*. <http://preval.org/files/Fetterman.pdf>, accessed 26 April 2013.
- Ghaffar A, IJsselmuiden C, Zicker F (eds). 2008. *Changing Mindsets: Research Capacity Strengthening in Low- and Middle-income Countries*. Geneva: COHRED, Global Forum for Health Research, and UNICEF/UNDP/World Bank/WHO Special Programme for Research and Training in Tropical Diseases (TDR). http://www.who.int/tdr/publications/documents/changing_mindsets.pdf, accessed 26 April 2013.
- Health Research for Action (HERA). 2007a. *Review of DANIDA-supported Health Research in Developing Countries: Main Report Volume One*. Reet: HERA. http://www.enrecahealth.dk/pdfs/20070309_DANIDA-Research_FINAL_report_volume_I.pdf, accessed 26 April 2013.
- HERA. 2007b. *Review of DANIDA-supported Health Research in Developing Countries: Annexes Volume Two*. Reet: HERA.
- House of Commons Science and Technology Committee (HCSTC). 2012. *Building Scientific Capacity for Development Fourth Report of Session 2012–13 HCC 377*. London: The Stationary Office Limited. <http://www.publications.parliament.uk/pa/cm201213/cmselect/cmsctech/377/377.pdf>, accessed 26 April 2013.
- Hovland I. 2007. *Making a Difference: MeOE of Policy Research (Working Paper 281)*. London: Overseas Development Institute. <http://www.odi.org.uk/sites/odi.org.uk/files/odi-assets/publications-opinion-files/2426.pdf>, accessed 26 April 2013.
- Keown K, Van Eerd D, Irvin E. 2008. Stakeholder engagement opportunities in systematic reviews: knowledge transfer for policy and practice. *Journal of Continuing Education in the Health Professions* **28**: 67–72.
- Kubisch AC, Auspos P, Brown P, Dewar T. 2010. *Voices from the Field III: Lessons and Challenges from Two Decades of Community Change Efforts*. Washington: Aspen Institute. http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/VoicesIII_FINAL_0.pdf, accessed 26 April 2013.
- Lincoln YS, Guba EG. 1985. *Naturalistic Inquiry*. Newbury Park: Sage Publications.
- Maessen O. 2005. *Intent to Build Capacity through Research Projects: An Examination of Project Objectives, Abstract and Appraisal Documents*. Ottawa: International Development Research Centre. http://www.idrc.ca/uploads/user-S/11635261571Intent_to_Build_Capacity_through_Research_Projects.pdf, accessed 26 April 2013.
- Mills A, Paimanzi R, Sewankambo N, Shasha W. 2010. *Mid-term Evaluation of the Consortium for National Health Research, Kenya*. Nairobi: Consortium for National Health Research [made available by the funders].
- Minja H, Nsanabana C. 2010. *TDR Research Capacity Strengthening Programmes: Impact Assessment Study Review for the Period 2000–2008 (Draft)*. Basel: Swiss Tropical & Public Health Institute (SwissTPH).
- Patton MQ. 1994. Developmental evaluation. *American Journal of Evaluation* **15**: 311–9.
- Patton MQ. 2011. *Developmental Evaluation: Applying Complexity Concepts to Enhance Innovation and Use*. New York: Guilford Press.
- Pawson R, Sridharan S. 2009. Theory-driven evaluation of public health programmes. In: Killoran A, Kelly M (eds). *Evidence-based Public Health: Effectiveness and Efficiency*. Oxford: Oxford University Press, pp. 43–61.
- Pederson JS, Chataway J, Marjanovic S. 2011. *CARTA (Consortium for Advanced Research Training in Africa): The Second Year*. Cambridge: RAND Europe [made available through the funder and contractors].
- Peebles D, Sachdeva N. 2008. *Gender Evaluation of Governance Equity and Health Program: Final Report on Phase 3—Gender Training Workshop*. Ottawa: International Development Research Centre.
- Podems D. 2010. *Evaluation of the Regional Initiative in Science and Education. The Carnegie-IAS Regional Initiative in Science and Education (RISE)*. Cape Town: OtherWise.
- Pope C, Ziebland S, Mays N. 2000. Qualitative research in health care: analysing qualitative data. *British Medical Journal* **320**: 114–6.
- Ransom J, Stackhouse J, Groenhout F, Geddes N, Darnbrough M. 2010. *Evaluating Commonwealth Scholarships in the United Kingdom: Assessing Impact in the Health Sector*. London: Commonwealth Scholarship Commission in the UK. <http://cscuk.dfid.gov.uk/wp-content/uploads/2011/03/evaluation-impact-health-report.pdf>, accessed 26 April 2013.
- Sachdeva N, Jimeno C, Peebles D. 2008a. *Gender Evaluation of Governance Equity and Health Program: Final Report on Phase I; Institutional Assessment*. Ottawa: International Development Research Centre.
- Sachdeva N, Peebles D. 2008. *Gender Evaluation of Governance Equity and Health Program: Final Report; Summary and Way Forward/Next Steps*. Ottawa: International Development Research Centre.
- Sachdeva N, Peebles D, Tezare K. 2008b. *Gender Evaluation of Governance Equity and Health Program: Final Report on Phase 2; Projects Review*. Ottawa: International Development Research Centre.
- Saini M, Shlonsky A. 2012. *Systematic Synthesis of Qualitative Research: A Pocket Guide for Social Work Research Methods*. New York: Oxford University Press.
- Smith E, Donovan S, Beresford P *et al.* 2009. Getting ready for user involvement in a systematic review. *Health Expectations* **12**: 197–208.
- Sridharan S, Campbell B, Zinzow H. 2006. Developing a stakeholder-driven timeline of change for evaluations of social programs. *American Journal of Evaluation* **27**: 148–62.
- Sridharan S, Nakaima A. 2011. Ten steps to making evaluations matter. *Evaluation and Program Planning* **34**: 135–46.
- Srivastava CV, Hodges A, Chelliah S, Zuckerman B. 2009. *Evaluation of the Global Research Initiative Program (GRIP) for New Foreign Investigators*. Bethesda: National Institutes of Health.

- van Velzen W, Kamal R, Meda N *et al.* 2009. *Independent External Evaluation Report of the European and Developing Countries Clinical Trials Partnership (EDCTP programme)*. The Hague: EDCTP. http://ec.europa.eu/research/health/infectious-diseases/poverty-diseases/doc/iee-report-edctp-programme_en.pdf, accessed 26 April 2013.
- Vasquez EE, Hirsch JS, Giang LM, Parker RG. 2013. Rethinking health research capacity strengthening. *Global Public Health*. Special Issue: Social and Political Dimensions of HIV and AIDS in Vietnam **8**(Suppl. 1): S104–24.
- Vogel I. 2011. *Research Capacity Strengthening: Learning from Experience*. London: UKCDS. http://www.ukcds.org.uk/_assets/file/features/UKCDS_Capacity_Building_Report_July_2012.pdf, accessed 26 April 2013.
- Vullings W, Meijer I. 2009. *NACCAP Mid term Review: Information Guide and Appendices*. Amsterdam: Technopolis Group.
- Whitworth JA, Kokwaro G, Kinyanjui S *et al.* 2008. Strengthening capacity for health research in Africa. *Lancet* **372**: 1590–3.
- Zuckerman B, Wilson A, Viola C, Lal B. 2006. *Evaluation of the Fogarty International Research Collaboration Awards (FIRCA) Program: Phase II Outcome Evaluation*. Bethesda: National Institutes of Health.